Predicting cell of origin from digitized images of hematoxylin and eosin-stained slides of diffuse large B-cell lymphomas using a cell-based deep-learning model

Anil Yüce,1 Jacob Gildenblat,2 Samaneh Abbassi-Sureshjani,1 Camille Laurent,3 Konstanty Korski3
1. Hoffmann-La Roche AG, Basel, Switzerland; 2. Department of Pathology, Rabin Medical Center, Petah Tikva, Israel; 3. Pathology Department CHU Toulouse IUCT Oncopole, CRCT, Université de Toulouse, Inserm, CNRS, Université Toulouse III-Paul Sabatier, Centre de Recherches en Cancérologie de Toulouse, Toulouse, France

Introduction
• Diffuse large B-cell lymphoma (DLBCL) can be classified by cell of origin (COO) as an activated B-cell-like (ABC) or germinal center B-cell-like (GCB) tumor.1
• ABC and GCB tumors carry differing prognoses; patients with ABC tumors generally have a higher risk of poor outcome than those with GCB tumors.2,3 • Timely and accurate classification and risk stratification is therefore important for appropriate diagnosis and patient care, as well as the future design of COO-based clinical trials.

Methods
• Algorithms were trained, validated and tested using data from the phase 2 CAVALLI (Clinical Trials.gov identifier: NCT02055820) and phase 3 GOYA (Clinical Trials.gov identifier: NCT12877411) trials.4,5 • The slides from GOYA were split into a training set and a test set for model tuning. The CAVALLI slides were used as an independent holdout set for further validation of the final model to prevent overfitting (Table 1).

Table 1. Clinical trial data sets used to train, validate and test the DL models

<table>
<thead>
<tr>
<th>Trial</th>
<th>ABC slides</th>
<th>GCB slides</th>
<th>Total slides</th>
<th>Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase 3 GOYA</td>
<td>106</td>
<td>202</td>
<td>308</td>
<td>Training</td>
</tr>
<tr>
<td>Phase 3 GOYA</td>
<td>21</td>
<td>46</td>
<td>67</td>
<td>Test</td>
</tr>
<tr>
<td>Phase 2 CAVALLI</td>
<td>57</td>
<td>110</td>
<td>167</td>
<td>Holdout</td>
</tr>
</tbody>
</table>

*These clinical trial data sets are publicly available.

Objective
• To develop and compare the performance of tile-level and nucleus-based DL models to perform COO classification using WSIs of H&E-stained pathology slides from patients with DLBCL.

Genome expression profiling was used to confirm the ground truth COO classification.

Following preprocessing, features were extracted using a custom fixed network pretrained on digital pathology images using the ResNet-50 neural network and the self-supervised Bootstrap Your Own Latent (BYOL) method. A multiple instance learning transformer was used to aggregate extracted features and predict the WSI label in a weakly supervised manner (Figure 1).

• Multiple instance and weakly supervised learning allow models to be developed using samples labeled at slide level, rather than requiring complete annotation of individual tiles, features and nuclei.
• For the tile-based model, 1024 features per tile were extracted from tumor regions of the WSIs.
• For the nucleus-based model, 256 features per nucleus were extracted from automatically segmented nuclei from the same tiles.

For the analysis of quantifiable feature sets, features were clustered by unsupervised k-means clustering (k = 16). Clusters were split into ABC- or GCB-significant groups according to the proportion of ground truth ABC and GCB slides, and quantifiable feature patterns were then analyzed for each cluster. Average quantifiable feature patterns were also calculated per tile and ABC- and GCB-labeled slides were compared using Student’s t-test.

An explainability model was applied, which assigned attention gradients to individual cells to reflect how strongly each cell influenced GCB or ABC classification. Morphological characteristics of cells that had been assigned positive (GCB) or negative (ABC) attention gradients were then inspected by a pathologist to determine the features that influenced COO classification.

Results
• Using the tile-based model, the areas under the receiver operating characteristic curves (AUCs) for the test set and holdout set were 0.73 and 0.67, and the average F1 scores were 0.68 and 0.61, respectively (Table 2).
• For the nucleus-based model, the AUCs for the test set and holdout set were 0.63 and 0.70, and the average F1 scores were 0.61 and 0.64, respectively (Table 2).
• On visual inspection of the highest-scoring tiles and nuclear patches, ABC-labeled tiles had higher tumor cell density than GCB-labeled tiles (Figure 2a) and ABC-labeled nuclei were larger than GCB-labeled nuclei (Figure 2b).

For quantifiable pathology features, ABC-significant clusters had a higher tumor cell density, lower lymphocyte density and higher mean cell area than GCB-significant clusters (Figure 2c).

The ratio of tumor to nontumor cells and the average tumor cell size were significantly higher for ABC-labeled slides (both p < 0.001).

Inspection of the explainability model attention gradients and nuclear features revealed that more ABC classifications than GCB classifications were related to bright areas of nuclei with lower chromatin content (Figure 3).

Table 2. Performance of the DL models

<table>
<thead>
<tr>
<th>Model type</th>
<th>Test set (21 ABC slides, 44 GCB slides)</th>
<th>Holdout set (67 ABC slides, 115 GCB slides)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC-labeled slide</td>
<td>ABC Activation</td>
<td>Average F1 score</td>
</tr>
<tr>
<td>Tile-based model</td>
<td>0.73</td>
<td>0.68</td>
</tr>
<tr>
<td>Nucleus-based model</td>
<td>0.63</td>
<td>0.61</td>
</tr>
</tbody>
</table>

*These clinical trial data sets are publicly available.

Figure 1. Workflow for predicting COO with tile- and cell-level morphology

Figure 2. Morphological features associated with ABC- or GCB-labeled WSIs

Figure 3. Explainability model attention gradients associated with ABC and GCB labeling

Conclusions
• These DL models demonstrated reasonable performance in COO classification of DLBCL from H&E-stained WSIs, without the use of ancillary techniques.
• The explainability model identified novel cellular and morphological features, including cell size and tumor cell density, that were associated with ABC or GCB classification.

Following further validation and regulatory approval, these models have the potential to supplement the diagnostic toolkit available to pathologists for reliable COO classification in the future.

Acknowledgments
The study was supported by Hoffmann-La Roche AG, Basel, Switzerland. Medical writing support was provided by Kay Schindl, PhD of Healthcare South, Ltd. The authors and Hoffmann-La Roche AG, Basel, Switzerland have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript associated with publication, practice, or research.

Disclosures
All authors declare no conflicts of interest. Financial disclosures include research grants (to the institution) and/or personal fees (to the authors) from Hoffmann-La Roche AG, Basel, Switzerland. Financial disclosures include research grants (to the institution) and/or personal fees (to the authors) from Genentech-DuPont Farmitalia, and personal fees (to the authors) from weighted and/or nonweighted research funding, from the following: Research grants from the National Health and Medical Research Council of Australia, the National Institutes of Health, and/or Johnson & Johnson. SA-S is an employee and stockholder of Hoffmann-La Roche AG, Basel, Switzerland. The other authors (AY, JG, and KL) have no relevant financial relationships to disclose.

References